Wind speed forecasting using Singular Systems Analysis

European Association for the Development of Renewable Energies, Environment and Power Quality (EA4EPQ)

International Conference on Renewable Energies and Power Quality (ICREPQ'13)
Bilbao (Spain), 20th to 22th March, 2013

Wind speed forecasting using Singular Systems Analysis

C. Skittides[1] and W.-G. Früh[1]

[1] Institute of Mechanical, Process and Energy Engineering, School of Physical Sciences and Engineering,
Heriot-Watt University, Riccarton, Edinburgh EH14 4AS (UK)
e-mail: cs261@hw.ac.uk

Abstract

We report on a method to predict wind speeds up to 24 hours ahead using a technique originating in Dynamical Systems and Chaos theory using a signal processing technique known as Singular Systems Analysis.

The method predicts wind speeds based on a set of previous measurements which were used to construct an attractor in an optimally defined phase space as a 'training set'. Current wind measurements can then be projected to onto that phase space to find most similar previous measurements. By tracing the evolution of these similar previous data, it is possible not only to forecast the wind speed but also to obtain a measure of the expected forecasting uncertainty.

The method was applied to a set of hourly wind speed data from a UK Meteorological Office weather station near Edinburgh. A comparison with a simple persistence prediction showed that the Singular Systems Analysis was both, consistently better at predicting wind speeds between 12 and 24 hours ahead than persistence, and also able to provide a meaningful forecasting uncertainty.

Key words

Wind resource, Wind speed forecasting, Statistical forecasting.

1. Introduction

Wind energy is one of the most established renewable energy forms. It has also the characteristic of a strongly intermittent form of energy with a large variability hence good wind resource assessement is of vital importance. The methods of analysis and prediction of wind behaviour are indeed of extreme importance for a good resource assessment.

Forecasting is an aspect of wind energy which has been under great investigation. It is associated with short-term prediction of wind speed. The forecasting horizons can be divided into the three following categories: 1) immediate-short-term (8 hours-ahead),2) Short-term (day-ahead) and 3) long-term (multiple-days-ahead) forecasting [1]. Several forecast models have been developed which can be categorized into physical, such as the Numerical Weather Prediction systems (NWPs), statistical, including linear methods such Auto Regressive Moving Average models (ARMA) or methods coming from artificial intelligence and machine learning fields such as Artificial Neural Networks (ANNs) [2] or even by hybrid approach methods which are a combination of statistical and physical methods with a use of weather forecasts and analysis of time series[1].

Factors such as the seasonality, time-of-day changes and weather systems are essential to be identified in terms of wind energy forecasting. The wind related data could be treated as dynamical systems so that cycles and random unusual behaviours that often characterise them can be identified, explained and understood. For example, for mean daily or hourly wind speed forecasts, i.e short-term horizon, the underlying atmospheric dynamics become of great importance. [3]. Thus the need of the creation of a tool that is capable of identifying trends, climate cycles and true outliers becomes vital.

Principal Component Analysis (PCA) is a statistical technique to identify dominant patterns of behaviour or response [4]. It is also known as Empirical Orthogonal Function (EOF) Analysis in the Meteorological and

Oceanographic community to identify the main circulation patterns in the Atmosphere and oceans, e.g. [5,6]. The application of the technique to time series is also known as Singular Systems Analysis (SSA) [7] which applies the PCA to a time series matrix generated from the measurements using Takens' method of delays [8].

This technique is now widely used for time series analysis of nonlinear dynamical systems in general and meteorology in particular, e.g. [6,9,10] as the analysis is very powerful to separate coherent dynamics from noise and it decomposes the measurements into an underlying invariant 'attractor' on which the dominant parts of the dynamics evolves, as well as a spectrum of the time-averaged contribution to the dynamics from the different attractor components

2. Forecasting algorithm

A *dynamical system* is used to model physical phenomena whose state (or instantaneous description) changes over time [11]. It is an approach to describe the behaviour over time of a system based on position and momentum in each direction, called a *phase space*. With complex systems one has to re-construct an equivalent to the phase space [7]. Other important definitions which involve dynamical systems are: the *phase space* which describes the system's variables, the *attractor* which defines the actual solution of the system and finally the *orbit* which is the path that the system follows during its evolution.

Furthermore, a method is needed to define equivalent variables to the ones of the phase space which is the *time-delay* method. It is a practical implementation of the dynamical systems since it aids in reconstructing the phase space of a dynamical system from an observed deterministic time series. The reconstruction of a phase space is indeed significant since it can extract useful information about the time series that characterises the system. Since the time-delay method is sensitive in the choice of the parameters that it uses for the analysis, Principal Component Analysis (PCA comes of use which can optimize phase space reconstruction. It is a non-parametric statistical method and by that is not limited to be of a certain distribution or linear relationship. PCA can separate noise from useful information applied to time-delay series [10]**.** It can also identify the number of needed time-delays and give a picture of their shape. Its goal is to explain important variability of the time series data and to extract useful information (i.e. hidden structures of the data) from its more relevant components in a reduced number of dimensions.

The mathematical procedure to carry out a PCA is through the Singular Value Decomposition (SVD) of the delay matrix. The three SVD/PCA outputs are the *singular values*, which measure total contribution of each dimension to total variance**,** the *singular vectors* which represent the dimensions and optimum phase space reconstruction and the *principal components* (PCs) which form an attractor and describe the system's time series and separate important variables from noise. In matrix notation, the singular value decomposition is written as

$$Y = P\Lambda S \tag{1}$$

where $Y(n,m)$ is the time-delay matrix with n the number of time points within time series and m the number of columns. If a delay matrix is constructed from the time series of a single variable, m is the number of delays. If a delay matrix is constructed from k different variables using m_w delays, then $m = km_w$. $P(n,m)$ is the principal component matrix, $\Lambda(m,m)$ is the diagonal matrix of singular values and $S(m,m)$ contains the eigenvectors.

The singular values represent a measure of the variance, more specifically the square root of the variance of the time series in corresponding dimensions and they can pick out the important variability of the data. The eigenvectors have the property of being orthonormal, i.e. orthogonal and of unit length and they span the dimensions of the phase space. They represent a measure of those dimensions that define a dynamical system, for instance they can replace position and momentum, two variables which can form a dynamical system.

When PCA is applied to the time-delay matrix, PC's are the time series of the coordinates of that trajectory in respect of these dimensions. PC's can replace the values of the position and momentum at any time. In more detail, this dynamical system's position of the reconstructed phase space can be given at any time precisely by position and momentum however when PCA is applied the PC's take over this role. Since there exists an eigenvalue matrix in PCA analysis it should be noted that both eigenvectors and PC's are normalised i.e. scaled to the amplitude of the dimensions used by PCA.

The training period has resulted in zero-shifting of the observable by, using the velocity as the example, its arithmetic mean, $U_m = \langle u \rangle$, and a scaling by its standard deviation, σ_u. The delay matrix is built up by choosing a sampling interval, τ, and a window length of M_w for the multi-variate time series from N_o observables or channels, each shifted to be centred around zero and scaled by their respective standard deviation. With a time series of length N_t, the delay matrix will have $N = N_t - \tau M_w$ rows and $M = N_o M_w$ columns with

$$Y^{i,j+(jo-1)M_w} = y_{j_o}\left(j+(i-1)\tau\right),$$

with the row index $i = 1 \ldots N$, the column index $j = 1 \ldots M$, and the observable index $j_o = 1 \ldots N_o$.

The key step in the analysis is to reconstruct an optimal attractor which separates signal from noise as much as possible. This is carried out by a singular value decomposition of the delay matrix,

$$Y^{N,M} = P^{N,M}\Lambda^{M,M}S^{M,M}, \tag{2}$$

where Y is the delay matrix from the measurements, P the Principal Components, Λ the diagonal matrix of

Singular Values, and S the Singular Vectors. This procedure is equivalent to an eigenvalue decomposition of the covariance matrix. Hence, the 'optimum' refers here to maximising the variance from the signals into a minimum number of orthonormal basis functions (EOFs). The average magnitude of contribution from each singular vector to the overall signal is measured by the singular value, and the principal components contain the time series (amplitudes) of the singular vectors.

By creating a reduced set of D_r principal components, singular values, and singular vectors, $P_r^{N,M_r}, \Lambda_r^{M_r,M_r}$ and $S_r^{M_r,M}$, respectively, one can produce a filtered time series of the original data by $Y_r^{N,M} = P_r \Lambda_r S_r$. Conversely, it is also possible to project a new time series onto that set of singular vectors by creating a delay matrix following the same procedure as for the training set, including using the mean and standard deviation from the data set to rescale the new data. This projection will then give principal components, P_n, to place the new data in this phase space

$$P_n = Y_n S_r^T \Lambda_r^{-1}. \tag{3}$$

To generate a single point in this phase space, the new time series must contain τM_w measurements. Conversely, if the new time series contains $\tau M_w + n_x - 1$ points, the projection results in a section of orbit containing n_x points.

The principle of the forecasting is to find similar records from the training periods, identified as the nearest neighbours to the current point or orbit section in phase space, and then follow how these neighbours evolved. The nearest neighbours are found by calculating the Euclidean distance between the new point, or the mean distance of each point of the section of orbit, to all other points or sections of the training attractor; for a single point: $d_i = \left|P_n - P_r^i\right|$ or for a section of orbit with n_x points: $d_i = \frac{1}{n_x}\sum_j \left|P_n^j - P_r^{i+j-1}\right|$. From this complete set of distances to all points of the training attractor, a limited number of nearest neighbours is selected, subject to a constraint that they do not come from adjacent points on the training orbit bur from different passes of the orbit through the neighbourhood. This can either be done by sorting all distances and rejecting those which come from adjacent points of the training time series, or by stepping through all distances, and skipping a set number of time points after having identified a local minimum of the distances. The number of nearest neighbours, n_n to use for the forecasting depends on the dimension of the reduced system and how densely the phase space is covered by the training attractor. If too few neighbours are chosen, the ensemble prediction might not capture the divergence or convergence of the attractor and hence not give a good estimate of the forecasting error. If too many neighbours are chosen, the nearest neighbours may not be that near and no longer be a good representation of the local dynamics, hence introducing errors into the forecasting.

Once the nearest neighbours have been identified, each can be moved forward in time by the forecasting horizon while sampling all intervening time steps. If entry k' of the training Principal components have been identified, then the entry $k = k' + n_x - 1$ is the neighbour to the latest measurement. A key assumption in the forecasting implicit here is that the current point will evolve alongside the past nearest neighbours, that is that the relative position of the state a time T in the future relative to the past nearest neighbour the same time interval T later will be identical to the relative position of the current measurement to the nearest neighbour. If the current distance vector to nearest neighbour j is $D_j = P_r^{k_j} - P_n^{n_x}$, then the prediction based on this nearest neighbour is $P_f^j(T) = P_r^{k_j+T} + D_j$. The ensemble of $P_f^j(T), j = 1 \dots n_n$ is then the ensemble prediction. One could then either calculate a mean prediction, and the change in distances to estimate the error growth, in the phase space, and then convert back to actual wind speeds, or one can convert the ensemble predictions into velocities, and then calculate the mean and error growth. Since it is more intuitive to collapse the ensemble in physical wind speed predictions than in phase space, this is the approach taken here. Each member of the ensemble is mapped back onto the delay matrix space by using $Y_f^j = P_f \Lambda_r S_r$. Each of the Y_f^j then returns the predicted wind speeds for the next T time steps as the entries $u_p^j(+1 \dots T) = Y_f^j(N - T + 1 \dots N, M_w)$.

This ensemble of predicted wind speeds can then be used to calculate the expected velocity as their average, and an estimate of the uncertainty based on the standard deviation: $\sigma_p(t) = \left\langle u_p^j(t)\right\rangle_j$.

3. Wind speed data

The data used for this analysis originated from the UK Met. Office – MIDAS Land Surface Stations [12] and more specifically, the station used was the Gogarbank surface station in Edinburgh, Scotland. The site used 10m high above ground anemometers and the data records used spanned from 1998-2010 with hourly mean wind readings stored to the nearest knot (1kn=0.5144m/s). For this analysis purposes wind speed and wind direction data were used with wind speed converted to m/s. Furthermore, the mean was removed from data and they were normalised by dividing with the standard deviation.

Regarding the forecasting analysis, the forecasting horizon used was from 1h to 24h ahead for hourly wind data measurements (τ=1) and the training periods used were 2008-2009 and 2000-2001 for different forecasting periods such as 2006, 2009 and 2010 depending on the training period used. Other variables such as the nearest neighbours n_n, reduced dimensions D_r, the overlap n_x, window length M_w and different forecasting years were also examined through PCA for the aforementioned models. Finally, in order to validate the results the error was calculated as the magnitude of the observed minus the forecasted data readings and the uncertainty was calculated from the standard deviation of the forecasted data.

Table 1. Forecasting models used in the PCA analysis

Training	n_x	n_n	D_r	M_w	Forecast	Forecasting horizon
2008-2009	1 **Range:** 1-3	5 **Range** 2-10	16 **Range:** 5-35	1 day **Range:** 1 day-2 weeks	**Main:** 2010 **Range:** 1999-2007	**Range:** 1-24h
2000-2001	1	10	30	2 days	**Range:** 2002-2009	**Range:** 1-24h

4. Results

A. Training

In terms of real wind data forecasting the following steps were undertaken. Initially, a phase space and attractor from the training period of Gogarbank wind speed data for 2008 and 2009 was constructed and truncated to 'important dimension' based on the singular values originating from the PCA results. As it can be seen from Figure 1, 20 singular values (D_r =20) seemed to be the leading ones and Figure 2 depicts a short section the original training period (green) alongside with the reconstruction of the truncated set (red).

After this, new data from Gogarbank for 2010 were acquired and as it can be seen in Figure 3 they were mapped onto the phase space from the training data (blue). Then the nearest neighbours (n_n), that is past events which were similar to the current wind were found (red). These neighbours were thus used to make an ensemble forecast (red lines), i.e. to follow how they evolved over time. In more detail, the blue point in Figure 3 comes from a week worth of hourly wind speed data (M_w=1 week) for Gogarbank 2010 and there can be seen that n_n=5 with n_x =1 where chosen for the prediction. In total, 60 different predictions were made from the training set of 2008 and 2009 for hourly data with weekly window with the forecasting horizons varying from 1 to 24 hours.

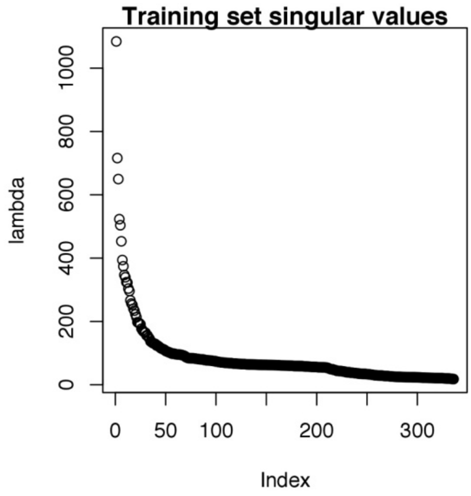

Figure 1. Singular values of PCA for training set

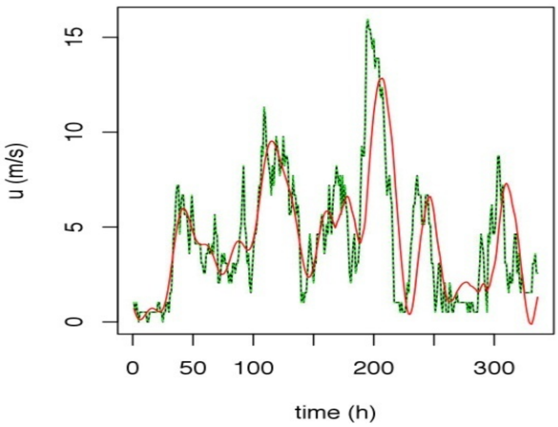

Figure 2. Original wind speed measurements and reconstruction from truncated attractor

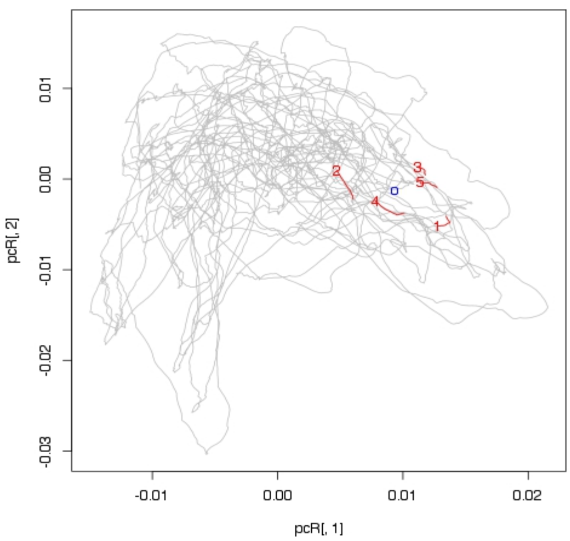

Figure 3. Attractor of training set with data mapped onto it.

B. Prediction

Figure 4 illustrates a validation of the ensemble forecast representing all 24 hours for one of the 60 prediction points. For this analysis purposes, the predictions made in the phase space were re-transformation to real wind speed and direction, and the mean prediction was calculated from the mean of the ensemble forecast (black circles). The forecasting uncertainty was also found with the use of the standard deviation (blue) and hence the comparison to the actual wind events was made (red lines). It can be therefore seen that there is growth or reduction of uncertainty over time which is consistent with the actual error. The estimate of error used in the analysis was the mean absolute error (MAE) which is a common measure for error used for forecasting purposes. It is of the form:

$$MAE(T) = \frac{1}{N_p}\sum_{i=1}^{N_p}\left|u_p\left(t_p+T\right) - u_{obs}\left(t_p+T\right)\right| \quad (4)$$

Where N_p is the number of forecasts made at times t_p for the wind speeds T hours ahead with u_p the prediction and u_{obs} the actual observed wind speed[13]. Moreover, the uncertainty was calculated as the standard deviation of the ensemble forecast (see function above).

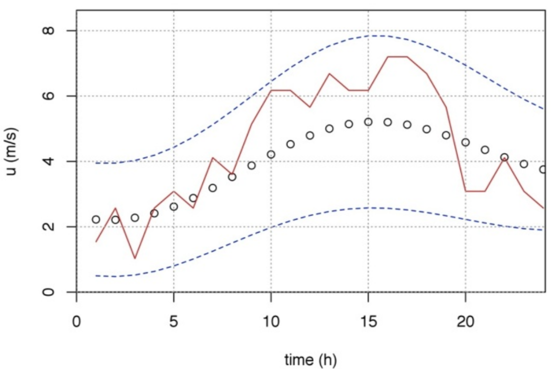

Figure 4. Comparison of actual, forecasted and uncertainty of wind speed

C. Comparison with other methods

The persistence method was then used in order to compare the PCA forecasting results. This method is simple and it just assumes that the wind speed from the starting point where it is calculated, it will remain unchanged for the rest of the forecasting horizon.

PCA filters out noise in the data however persistence only accounts for what has just happened following a random event. Hence by combining the two, we achieve slower dynamics of the PCA. After performing this comparison and applying several inputs for the different parameters used by PCA, it was concluded that adding a filter to the dataset would improve the results. This filter is the prediction minus the initial error counted and applied from the 5th hour of the forecasting horizon up to the 12th. From the 5th hour when the filter was applied and up to the 12th horizon hour, the error subtracted was reduced in quantity. The filter is of the form:

$$u_{f,i} = u_{PCA,0} - \left(u_{PCA,0} - u_0\right)\frac{N_f - i}{N_f}; i = 0,..,N_f \quad (5)$$

or

$$u_{f,i} = u_{PCA,i}; i > N_f \quad (6)$$

where i = the ith step ahead in the forecast horizon , N_f is the filter length, u_{PCA,N_f} is the ensemble forecast and u_0 is the initial error.

Figures 5 illustrates the averaged error and uncertainty growth of the actual readings, PCA and persistence results. The red line corresponds to the distance of the red line (actual readings) minus the circles (PCA results) of Figure 4. The red line thus should ideally be below the green line (persistence method) since this indicates that the error is smaller when using PCA in comparison with the persistence method. From Figure 5 it can be clearly seen that adding the filter aids in achieving this.

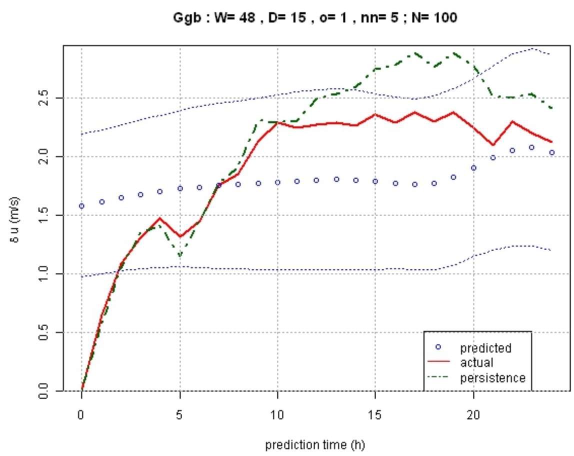

Figure 5. Error growth with filtered data

D. Sensitivity to parameters

In order to validate the aforementioned analysis different models have been attempted with different entries of variables used for the PCA analysis as shown in Table 1. Some representative results of the undertaken analysis are shown in this chapter. Since in the previous section it was concluded that the use of filter in the data was improving the PCA results, all the analyzed models included the filtered data instead. A performance index percentage was thus introduced to examine the improvement of PCA in comparison with the persistence method. This index is calculated as the average over the range of prediction times of the difference between the mean persistence and PCA forecasting errors, divided by the respective average of the mean persistence error multiplied by 100.The error measure as it was explained in the previous section, was MAE.

More specifically, Figures 6 and 7 show the performance index of the results for the different entries of D_r , and n_n . Figure 6 indicates the reduced dimensions improvement. It can be seen that different choice of reduced dimensions results in a big variation of the percentage of improvement. The amount of dimensions which the improvement seems to be more consistently big for (5.6%) is around 16. It should be noted that adding more dimensions results in adding more information but whether this information is useful or not is another issue which should be of further investigation and of course depends on the site and wind dynamics used for the analysis.

Finally, Figure 7 indicates that choosing 5 nearest neighbours seems to result in the best improvement, again around 11.2%. Using too few or too many neighbours might not be appropriate since with too few (i.e. less than 5) the information we use for the analysis might me too little whereas on the contrary, using too many (i.e. more than 5) might initially show that we can obtain more information however these neighbours might actually lie very far apart from each other in the phase space.

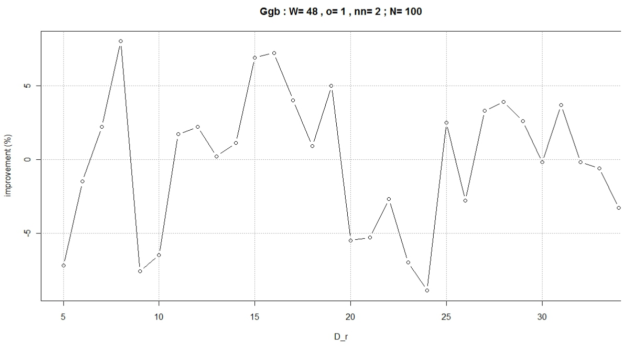

Figure 6. Performance index for different embedding dimension

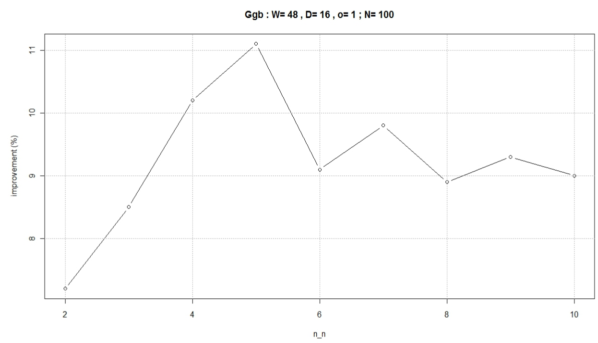

Figure 7. Performance index for different nearest neighbours

5. Discussion and Conclusions

The main conclusions of this research that can be made are firstly that PCA is capable of identifying weather cycles and a dynamical link between two sites, reference and target, that form an attractor. Furthermore, it was found that it can be used for wind forecasting several hours ahead and also it can obtain a measure of this forecasting uncertainty. Hence, it has clear potential to be used for MCP-type resource assessment as well as for operational wind power forecasting. It was also found that with the use of filtering, PCA outperformed the persistence method. Finally, testing the PCA performance with sensitivity analysis, it was found that the dimensions and nearest neighbours used play an important role in the PCA results.

Thus the future work that will be carried out should be focused on first, to validate fully the forecasting attempt with the use of other training and prediction periods but also by using other PCA parameter choices such as wind direction or temperature etc. Other types of data such as wind farm and Met office data should be also used. Extending the forecasting methodology for MCP methods would be the following step but of course challenges such as mapping data from one site onto the appropriate place of the phase space based on both sites will arise.

Acknowledgements

The authors would like to thank the British Atmospheric Data Centre (BADC) and SgurrEnergy Ltd. for their help.

References

[1] Wang X, Guo P, Huang X. A Review of Wind Power Forecasting Models. Energy Procedia 2011; 12:770-778.

[2] Erdem E, Shi J. ARMA based approaches for forecasting the tuple of wind speed and direction. Applied Energy 2011; 88: 1405-1014.

[3] Landberg L. Short-term prediction of the local wind conditions. Journal of wind engineering and Industrial Aerodynamics 2001; 89 (3-4): 235-245.

[4] Everitt B and Dunn G. *Applied Multivariate Data Analysis*. Arnold, 2nd edition, 2001.

[5] Palmer, TN. A nonlinear dynamical perspective on climate change. *Weather* 1993; 48:314–326.

[6] Bakalian F, Ritchie H, Thompson K, and Merryfield W. Exploring atmosphere-ocean coupling using Principal Component and Redundancy Analysis. *Journal of Climate* 2010; 23(18):4926–4943.

[7] Broomhead DS, Jones R, King, JP, Pike ER. Singular system analysis with application to dynamical systems in ER Pike and LA Lugiato (eds.). *Chaos, noise and fractals*. Adam Hilger, 1987; 15-27.

[8] Takens F. Detecting strange attractors in turbulence. In Rand DA and Young LS, editors, *Dynamical systems and turbulence*, Warwick 1980: *Lecturenotes in Mathematics* 898, 366–381. Springer-Verlag, Berlin, 1981.

[9] Allen MR, Smith LA. Investigating the origins and significance of low-frequency models of climate variability. *Geophysical Research Letters* 1994; 10: 883-886.

[10] Read PL. Phase portrait reconstruction using multivariate singular systems analysis. *Physica D* 1993; 69: 353-365.

[11] Brown University: A tutorial on learning dynamical systems, http://www.cs.brown.edu/research/ai/dynamics/tutorial/Documents/DynamicalSystems.html (accessed 16 April 2012).

[12] British Atmospheric Data Centre. *Met Office – MIDAS Land Surface Station data*. http://www.badc.ac.uk (accessed 16 April 2012).

[13] Shi J, Guo J, Zheng S. Evaluation of hybrid forecasting approaches for wind speed and power generation time series. Renewable and sustainable Energy Reviews 2012; 16: 3471-3480.